Carved Visual Hulls for High-Accuracy Image-Based Modeling

Yasutaka Furukawa (yfurukaw@uiuc.edu) and Jean Ponce (ponce@cs.uiuc.edu)
Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

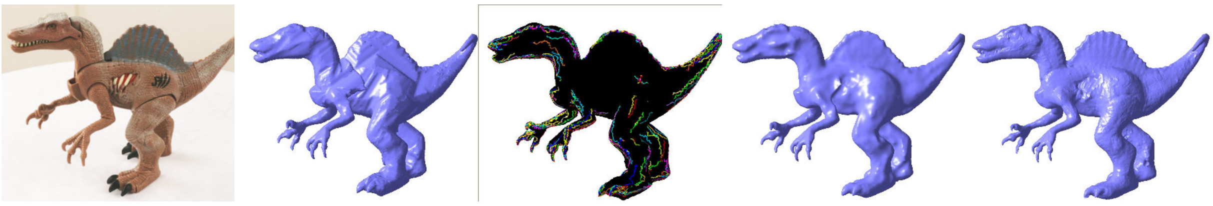

Figure 1: Constructing the carved visual hull of a toy dinosaur. From left to right: one of the 24 input photographs, the raw visual hull, the rims found on its surface, the carved visual hull after graph cuts, and the final 3D model after iterative refinement.

The relative accuracy of high-end laser range scanners can be as high as 1/10,000, allowing the construction of high-accuracy solid models of complex shapes from registered depth maps [Levoy et al. 2000]. Comparable accuracy levels can be achieved using "ordinary" cameras and sophisticated photogrammetric methods, but these typically output a relatively sparse set of points and require markers [Uffenkamp 1993]. Computer vision approaches to image-based modeling from calibrated photographs construct solid object models and do not need markers [Baumgart 1974; Kutulakos and Seitz 2000], but their relative accuracy is typically below 1/200. [Hernandez and Schmitt 2004] propose to use the visual hull [Baumgart 1974] to initialize the deformation of a surface mesh under the influence of rim- and photo-consistency constraints expressed by gradient flow forces (see [Keriven 2002] for a related approach). Although this method yields excellent results, its reliance on iterative refinement makes it susceptible to local minima. To overcome this problem, we propose a combination of global and local optimization techniques to carve the surface of the visual hull. The algorithm proposed in [Lazebnik 2002] is first used to construct a combinatorial mesh representation of the visual hull surface in terms of polyhedral cone *strips* meeting at *frontier points* where two visual rays are tangent to the surface (Figure 2). Photo-consistency constraints are then used to refine this initial surface in three consecutive steps: (1) dynamic programming is used to identify the *rims* where the surface grazes the visual hull as the most photo-consistent paths between successive frontier points in the corresponding strips; (2) with the rims now fixed, the visual hull is carved by graph cuts [Boykov and Kolmogorov 2003] to globally minimize the image discrepancy of the surface and recover the main surface features, including concavities—which, unlike convex and saddle-shaped parts of the surface, are not captured by the visual hull; and (3) iterative (local) refinement [Keriven 2002] is finally used to recover surface detail.

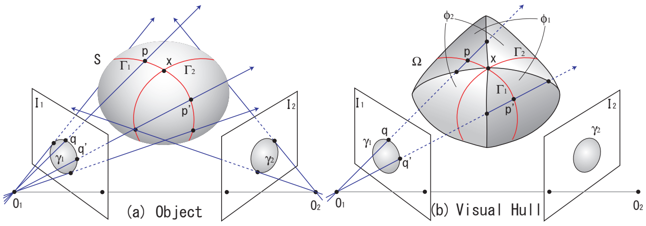

Figure 2: Visual hulls, cone strips and rims: (a) an egg-shaped object is viewed by 2 cameras with optical centers O_1 and O_2; the point x is a frontier point; (b) its visual hull is constructed from two apparent contours γ_1 and γ_2, the surface Ω of the visual hull consisting of two cone strips ϕ_1 and ϕ_2.

Figure 1 illustrates this process with the construction of the carved visual of a toy dinosaur from 24 calibrated photographs, with a mean projected object size of about 4Mpixel, and contours extracted interactively to sub-pixel precision. The toy dinosaur is about 20cm in diameter, with fine surface details including fin undulations, and scales in the neck and near the fin. These details are well captured by the model, even though the corresponding height variations are well below 1mm. Similar levels of detail have been obtained with several other real data sets, including the skulls of three prehistoric and modern human specimens, an action figure, and a person. Next on our agenda are (a) incorporating wide-baseline matching techniques to improve the reconstruction of deep concavities, and (b) applying bundle adjustment [Uffenkamp 1993] to simultaneously refine both the camera parameters and the surface shape, with the goal of achieving at least a tenfold increase in relative accuracy.

Acknowledgments. This research was partially supported by the National Science Foundation under grant IIS-0312438.

References

BAUMGART, B. 1974. *Geometric modeling for computer vision*. PhD thesis, Stanford University.

BOYKOV, Y., AND KOLMOGOROV, V. 2003. Computing geodesics and minimal surfaces via graph cuts. In *ICCV*, 26–33.

HERNANDEZ, C., AND SCHMITT, F. 2004. Silhouette and stereo fusion for 3D object modeling. *CVIU 96*, 3, 367–392.

KERIVEN, R. 2002. A variational framework to shape from contours. Tech. Rep. 2002-221, ENPC.

KUTULAKOS, K., AND SEITZ, S. 2000. A theory of shape by space carving. *IJCV 38*, 3, 199–218.

LAZEBNIK, S. 2002. Projective visual hulls. Tech. Rep. MS Thesis, UIUC.

LEVOY, M., PULLI, K., CURLESS, B., RUSINKIEWICZ, S., KOLLER, D., PEREIRA, L., GINZTON, M., ANDERSON, S., DAVIS, J., GINSBERG, J., SHADE, J., AND FULK, D. 2000. The digital michelangelo project: 3D scanning of large statues. In *SIGGRAPH*, 131–144.

UFFENKAMP, V. 1993. State of the art of high precision industrial photogrammetry. In *Third International Workshop on Accelerator Alignment*.